Understanding Society

Working Paper Series

No. 2024 – 09

September 2024

Measuring expenditure with a mobile app: Do probability-based and nonprobability panels differ?

Annette Jäckle (University of Essex)

Carina Cornesse (German Institute for Economic Research & University of Bremen)

Alexander Wenz (University of Mannheim)

Mick P. Couper (University of Michigan)

Understanding Society
THE UK HOUSEHOLD LONGITUDINAL STUDY

Economic
and Social
Research Council

Non-technical summary

We examine a novel aspect of data collected in probability and nonprobability panels, by comparing data collected with a mobile app: respondents in each panel were invited to participate in a month-long study, reporting all their daily expenditures in the app. In line with most of the research on nonprobability and probability-based panel data, our results indicate that the differences in the nonprobability and probability-based panel recruitment processes lead to differences in the data gathered (e.g., socio-demographic characteristics, financial behaviour, digital skills and behaviour, and substantive outcomes of interest) from these data sources. Our findings also show that these differences in the data are difficult to eliminate by weighting. The only data quality aspect for which we did not find evidence of differences between the nonprobability and probability-based panel was behaviour in using the spending app. This finding is contrary to the argument that nonprobability online panel participants try to maximize their monetary incentive at the expense of data quality. However, this finding is in line with some of the scarce existing literature on response behaviour in surveys which is inconclusive regarding the question of whether nonprobability online panel participants answer questions less conscientiously than probability-based panel respondents.

Measuring expenditure with a mobile app: Do probability-based and nonprobability panels differ?

Annette Jäckle (University of Essex)
Carina Cornesse (German Institute for Economic Research & University of Bremen)
Alexander Wenz (University of Mannheim)
Mick P. Couper (University of Michigan)

Abstract: Previous studies have shown that, while nonprobability panels are much cheaper to maintain than probability-based panels, estimates are often less accurate, even when taking account of differences in socio-demographic panel sample composition. In this paper we examine expenditure data collected with a mobile app over a period of one month in a probability-based and a nonprobability panel. We find differences between the app study samples in who participates in the mobile app study and in the expenditure captured with the app, even after accounting for differences in panel sample composition, but no differences in how participants used the app.

Keywords: spending diary, online panel, mobile app data, probability sample, nonprobability sample

JEL classification: C81, C83

Acknowledgements: This research was funded by grants from the UK Economic and Social Research Council (ESRC) and the National Centre for Research Methods (NCRM) ES/N006534/1and by ESRC funding for the Understanding Society survey (ES/N00812X/1). This work was also supported by the Collaborative Research Center SFB 884 "Political Economy of Reforms" (projects A8 and Z1), funded by the German Research Foundation (DFG). This study design and analysis was not preregistered.

Corresponding author: Annette Jäckle, Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex CO7 9FJ, United Kingdom, aejack@essex.ac.uk.

1. Introduction

Many researchers use nonprobability online panels to collect large amounts of survey data relatively quickly at low cost. However, nonprobability online panels rely on volunteers who often participate in multiple panels (Hillygus, Jackson, and Young 2014) and who are mainly motivated by the monetary compensation (Keusch, Batinic, and Mayerhofer 2014). They are commonly criticised for their poor performance in accurately representing the general population (MacInnis, Krosnick, Ho, and Cho 2018). However, probability-based panels face data quality issues, too, and some have argued that ever-decreasing survey response rates make probability-based panels indistinguishable from nonprobability online panels in terms of data quality (Wang, Rothschild, Goel, and Gelman 2015; Gelman, Goel, Rothschild, and Wang 2016). Nevertheless, previous studies have shown that probability-based panels produce more accurate estimates than nonprobability online panels (see Cornesse et al. 2020 for an overview).

So far, research on comparing results from probability and nonprobability online panels has focused on questionnaire-based survey data. In this paper, we compare these panels on a new dimension: we examine what happens when panel members are asked to use a mobile app to record their spending every day for a month. We use data from a diary study on financial spending that was implemented in parallel in the *Understanding Society* Innovation Panel, a probability-based mixed-mode panel of households in Great Britain, and in a commercial nonprobability online panel. Participants were asked to install an app on their mobile device and use it to report all spending over a period of one month. We compare the app study data collected in the two panels to answer the following research questions:

RQ1: Do the app study participants in the probability-based panel have different characteristics than those in the nonprobability online panel?

RQ2: Are there differences between the panels in how participants use the app?

RQ3: Are there differences between the panels in expenditure estimates, i.e., the study's main outcome of interest?

RQ4: Do the differences between the panels in expenditure estimates remain after weighting?

2. Conceptual Framework: Differences Between Probability-Based and Nonprobability Panels

There are two types of 'panels': 1) 'traditional' panels, where participants are surveyed on a core set of topics repeatedly, typically at less frequent intervals, to provide longitudinal data and 2) 'access panels', where participants are surveyed on a variety of different topics, often at frequent intervals. Both types of panels can be recruited using probability or nonprobability sampling methods, or a combination of both (Callegaro et al. 2014). Examples of traditional probability panels include the U.S. Panel Study of Income Dynamics, the UK Household Longitudinal Study, and the German Socio-Economic Panel. Nonprobability access panels are typically online panels run by commercial companies, such as Lightspeed, YouGov, Toluna, and OnePoll. In addition, there are probability-based online panels that combine frequent data collection on different topics with a traditional panel element collecting longitudinal data. Examples include the Understanding America Study and AmeriSpeak in the U.S., the German Internet Panel, and the Dutch LISS panel.

The researcher's choice of the sample recruitment procedure has consequences for the resulting participant samples and can, thereby, influence the estimates calculated on the basis of the gathered data (Mercer, Kreuter, Keeter, and Stuart 2017). To recruit a probability-based panel, researchers draw a random sample of units (individuals or households) from a sampling frame, such as an address list or population register. Sampled units are then approached with a request to participate in the panel. Because the probability sampling process is usually time consuming and expensive, great effort is often expended to gain contact with the sampled units (e.g., multiple contact attempts) and establish cooperation with the request to participate in the panel (e.g., by offering multiple modes of data collection). Participants are usually asked to participate in infrequent (often annual) surveys for a single sponsor (e.g., a university) and compensated with monetary incentives.

To recruit a nonprobability online panel, researchers usually disseminate open invitations, whether via online advertisement, by sending invitation emails via newsgroups and mailing lists, or by placing a panel invitation question at the end of a pop-up survey (see Callegaro et al. 2014 for an overview of nonprobability online panel recruitment methods). Recently, it has also become popular to recruit nonprobability panels based on interactive features implemented in online media articles or via social media (see Zindel 2022 for an overview). The likelihood and frequency of exposure to the open invitations depends on whether people have a chance of being exposed (e.g., whether they have an internet connection, whether and how often they visit any of the websites that display a particular ad, or whether they are enrolled in any mailing list over which an invitation is disseminated). The likelihood that a person who is exposed to an open invitation becomes aware of this invitation depends on whether the invitation catches their eye (e.g., whether a banner ad uses bright colours, or an invitation email header contains a promise of attractive incentives). The likelihood that a person who is aware of an open invitation to volunteer then joins the nonprobability online panel depends on the person's time constraints, topic interest, motivation, internet access constraints, and skills (e.g., digital literacy). Once a pool of people has volunteered to participate in a nonprobability online panel, researchers select panel members for a particular study. Sometimes this is done using quota sampling to achieve some balance with regard to a limited number of characteristics, such as age, gender, and geographic region. Nonprobability online panel members are typically exposed to a large number of survey requests (often several times a month or more frequently) for different sponsors or clients and are compensated with small monetary or in-kind incentives (e.g., points) for each completed survey.

In addition to differences in selection and recruitment methods, probability-based and nonprobability panels differ on many other dimensions, including how much effort is put into retaining sample members (i.e., reducing panel attrition), the degree of panel maintenance (e.g., removing fraudulent responses or dropping inattentive respondents), the frequency, type and size of survey requests, and the range of topics covered. These design differences may all lead to differences in sample composition, motivation, and interest of sample persons to participate in specific tasks or activities.

Overall, we expect the differences in the probability-based and nonprobability panel recruitment and retention processes to affect who is selected into the respective panels, which, in turn, might lead to differences in app study data gathered on these panels.

3. Expectations and evidence from previous empirical studies

A number of studies have examined whether it is sufficient to use samples from nonprobability online panels rather than investing in probability-based methods of survey data collection. For this purpose, researchers typically implement a questionnaire with identical survey questions, answer options, and fieldwork periods, among other design features kept identical, in at least one probability-based survey and at least one nonprobability online panel.

Most studies of this type focus on assessing how accurately probability-based and nonprobability surveys represent the intended target population. Some of the studies compare nonprobability online panels to probability-based face-to-face (Malhotra and Krosnick 2007; Loosveldt and Sonck 2008; Szolnoki and Hoffmann 2013; Ansolabehere and Rivers 2013; Dutwin and Buskirk 2017; Sturgis et al. 2018; Dassonneville, Blais, Hooghe, Deschouwer 2020) or telephone surveys (Szolnoki and Hoffmann 2013; Ansolabehere and Schaffner 2014; Gittelman, Thomas, Lavrakas, and Lange 2015; Pasek 2016; Dutwin and Buskirk 2017; Sohlberg, Gilljam, and Martinsson 2017; Pennay, Neiger, Lavrakas, and Borg, 2018; Legleye et al. 2018). Other studies compare nonprobability online panels to probability-based online panels (Chan and Ambrose 2011; Steinmetz, Bianchi, Tijdens, and Biffignandi 2014). A number of studies also try to disentangle potential mode and sampling effects by comparing samples from nonprobability online panels to probability-based offline surveys as well as probability-based online panels (Berrens, Bohara, Jenkins-Smith, Silva, and Weimer 2003; Chang and Krosnick 2009; Yeager et al. 2011, Scherpenzeel and Bethlehem 2011; Kennedy et al. 2016; Brüggen, van den Brakel, and Krosnick, 2016; MacInnis et al. 2018). Furthermore, some studies examine whether weighting adjustments improve the accuracy of nonprobability online panels using a variety of weighting procedures, such as raking (Berrens et al. 2003; Chang and Krosnick 2009; Pasek 2016; Dutwin and Buskirk 2017; Sturgis et al. 2018), poststratification (Loosveldt and Sonck 2008; Yeager et al. 2011; Gittelman et al. 2015; MacInnis et al. 2018; Pennay et al. 2018), or propensity weighting (Berrens et al. 2003; Loosveldt and Sonck 2008; Steinmetz et al. 2014; Pasek 2016; Dutwin and Buskirk 2017; Sturgis et al. 2018). Little attention is usually paid to differences in study participation behaviour between probability-based surveys and nonprobability online panels (for notable exceptions see Chang and Krosnick 2009, Greszki, Meyer, and Schoen 2014, and Cornesse and Blom 2020).

While the existing literature compares probability-based surveys to nonprobability online panels, our study goes beyond survey data and instead focuses on an additional task that people are asked to do: installing and using a mobile app to report their spending over a month. In the following, we discuss relevant existing evidence from the survey literature and describe our expectations regarding the mobile app study.

RQ1: Do the app study participants in the probability-based panel have different characteristics than those in the nonprobability online panel?

Previous studies have focused on comparing the composition of probability-based surveys and nonprobability online panels by deriving aggregate indices, such as the absolute average bias (Kennedy et al. 2016), the largest absolute error (Yeager et al. 2011), or the root mean squared error (MacInnis et al. 2018). Nonprobability online panels have repeatedly been found to be more selective in that they deviate more from population benchmarks than probability-based surveys (Cornesse et al. 2020).

Given the focus on aggregate indices in the existing literature, it is difficult to identify common patterns at the variable level. However, it seems that in nonprobability online panels, there is often a stronger overrepresentation of people who are middle-aged (Malhotra and Krosnick 2007; Legleye et al. 2018; Dassonneville et al. 2020), female (Malhotra and Krosnick 2007; Legleye et al. 2018; Dassonneville et al. 2020), and highly educated (Malhotra and Krosnick 2007; Legleye et al. 2018; Dassonneville et al. 2020; MacInnis et al. 2018) compared to probability-based surveys. In our study, we, therefore, expect that app study participants from the nonprobability online panel are more likely to be middle-aged, female, and highly educated than those from the probability-based mixed-mode panel (H1.1).

Since nonprobability online panel recruitment usually relies on open invitations disseminated via the internet and on volunteering rather than random selection, nonprobability online panels are also likely to systematically exclude or misrepresent some people beyond primary socio-demographic characteristics. For example, people without internet access do not have a chance of being exposed to invitations and people who are frequently online have a particularly high chance of exposure to such invitations. In addition, people who consider the costs of participation to be low (e.g., because they have high levels of digital skills) and the benefits of participation to be high (e.g., because the promised monetary or in-kind incentives appeal to them) are more likely to volunteer than people who consider the costs to be high and the benefits to be low. We, therefore, expect participants from the nonprobability panel to have higher levels of digital skills and greater interest in their finances than participants from the probability-based mixed-mode panel. Having higher levels of digital skills includes greater experience with and confidence in using digital devices (H1.2) as well as being more willing and less concerned to participate in additional smartphone-based tasks, such as taking photos or allowing GPS tracking (H1.3). Having greater interest in their finances includes being more involved in behaviours related to finances, such as budget keeping and bank balance checking (H1.4).

RQ2: Are there differences between the panels in how participants use the app?

Keusch et al. (2014) found that for 40% of the newly recruited members of a nonprobability online panel the monetary incentives were an important motive for joining the panel. In addition, this monetary motivation was the strongest predictor of actual participation in panel survey waves as compared to other motives such as curiosity, entertainment, or novelty. Similarly, Sparrow (2006) found that 52% of newly recruited nonprobability online panel members stated that they participated in the panel because of the monetary incentive provided. Other reasons for participating, such as survey enjoyment and interest in the survey topic, were selected significantly less often. While the studies by Keusch et al. (2014) and Sparrow (2006) suggest that nonprobability online panel participants are mainly motivated by the promise of monetary incentives, the same does not seem to apply to probability-based panel participants. For example, in the LISS panel only a minority (15.2%) of panel participants said that their most important motive for participating in the study was the financial rewards while most participants said that their main reason for participating was either that they think it is important to contribute to science (16.4%), contribute to society (13.6%), or help the researchers (13.0%; numbers based on own calculations, data retrieved from www.lissdata.nl).

Research on the consequences of the incentive-oriented motivation of nonprobability online panel members for survey data quality is scarce and results are mixed. On the one hand, Cornesse and Blom (2020) found that straight-lining in grid questions was significantly more likely in seven nonprobability online panels than in three probability-based online panels. On the other hand, Chang and Krosnick (2009) found the opposite when comparing a nonprobability online panel with a probability-based online panel. Similarly, Greszki et al. (2014) found that a nonprobability online panel performed worse in terms of survey response speed than a probability-based online panel, while Chang and Krosnick (2009) found that a nonprobability online panel performed better in terms of random measurement error across multiple measures of the same construct than a probability-based telephone survey. Finally, Cornesse and Blom (2020) did not find any generalisable difference between nonprobability online panels and probability-based online panels with regard to survey item nonresponse or midpoint selection.

In line with the general finding in the literature that many nonprobability online panel members are motivated to volunteer to join the panel primarily for monetary reasons, we expect that app study participants from the nonprobability sample are more likely to apply strategies to maximise their monetary compensation (H2.1) and to minimise their effort (H2.2) than app study participants from the probability-based sample.

RQ3: Are there differences between the panels in expenditure estimates, i.e., the study's main outcome of interest?

Generally, many studies find that differences between probability-based and nonprobability panels matter for a diverse set of substantive outcomes, such as voting behaviour (Malhotra and Krosnick 2007; Chang and Krosnick 2009; Sturgis et al. 2018), health behaviour (Yeager et al. 2011), consumption behaviour (Szolnoki and Hoffmann 2013), sexual behaviour and attitudes (Erens et al. 2014; Legleye et al. 2018), and political attitudes (Malhotra and Krosnick 2007; Loosveldt and Sonck 2008).

In our study we similarly expect differences in participant characteristics and study participation behaviour to result in differences between samples in key outcomes. In particular, if women, middle-aged people, and highly educated people are more likely to participate in a nonprobability online panel than a probability mixed-mode panel (H1.1), this should result in differences in the distribution of expenditure across spending categories.

In addition, if app study participants from a nonprobability online panel maximise their incentive-to-effort-ratio by reporting fewer spending events than those from probability-based online panels (H2.1, H2.2), this should result in lower expenditure estimates. Compared to estimates from a probability-based panel, we therefore expect that estimates from the nonprobability panel suggest lower average expenditures, for example on eating out and other leisure activities (H3).

RQ4: Do the differences between the panels in expenditure estimates remain after weighting?

Most previous studies conclude that significant differences between probability and nonprobability samples remain after weighting (Schonlau et al. 2004; Duffy, Smith, Terhanian, and Bremer 2005; Schonlau, van Soest, and Kapteyn 2007; Schonlau, van Soest, Kapteyn, and Couper 2009; Smyk, Tyrowicz, and van der Velde 2021; Mercer et al. 2017). Furthermore, some studies conclude that differences remain even after adding non-demographic characteristics to the weighting schemes (Lee 2006; Dutwin and Buskirk 2017; Mercer et al. 2017).

In our study we expect that if the probability and nonprobability samples differ in participant characteristics, study participation behaviour, and substantive outcomes, it is unlikely that these differences all vanish after applying common weighting procedures based on socio-demographic characteristics. Including additional variables is likely to further reduce the differences between probability and nonprobability samples if the added variables are related to the nonprobability sample recruitment process and/or the study data of interest. For example, if nonprobability sample members use the internet more frequently and with greater confidence than probability sample members, adding variables related to digital skills might help reduce differences in internet usage behaviour. We therefore expect that differences between probability and nonprobability samples will decrease when weighting for socio-demographic characteristics (H4.1), decrease further when adding financial behaviours (H4.2), and decrease further still when adding digital affinity measures related to the app use task (H4.3).

4. Data

Spending Study 2 (SS2; University of Essex, Institute for Social and Economic Research 2022) was implemented in May to December 2018, using two different samples in Great Britain: a probability panel (the *Understanding Society* Innovation Panel or IP) and a nonprobability panel (the Lightspeed UK online access panel, or AP). The samples are described below. This was a follow-up to an earlier study (Spending Study 1), carried out in 2016. The first study was only implemented in the probability-based IP and therefore is not included in the present analyses. See Jäckle et al. (2018) for details of the first study.

Participants were asked to download a mobile app and use it for one month to report their spending. The same app was used in both samples and was compatible with iOS smartphones and Android smartphones and tablets. The design and functionality of the app were based on findings from qualitative interviews with members of the general public about how the app could best support participants in reporting their daily expenditure (Suffield et al. 2018). Participants were asked to use the app to record all direct debits and standing orders that would automatically come out of their bank accounts during the month. In addition, they were asked to use the app every day to report all purchases, by selecting a purchase category and then entering the value of the purchase. On days on which they did not spend any money they were asked to report 'no purchases today' in the app. Sample members who did not use the app were invited to use a browser-based version instead. However, as take up of the browser-based version was very low in the IP, this paper focuses on participants in both samples who used the app.

Participants were told that they could earn incentives for every day on which they used the app at least once (including to report that they had not made any purchases that day), a bonus if they used the app every day, and incentives conditional on completing the direct debit / standing order section and a debrief questionnaire at the end of the study. In the IP, the daily incentive was £0.50, the bonus for completing the month was £10, the incentives for the direct debit section £1 and for the debrief questionnaire £3. In total participants could earn up to £29.50. For the AP, Lightspeed UK administered the incentives according to their standard rewards policy: panellists could earn a maximum of 500 points (equivalent to about £5) and could exchange their incentives for vouchers or charity donations. For details of SS2, including screenshots of the app, see Jäckle et al. (2019).

4.1 The Understanding Society Innovation Panel (IP)

The IP is a stratified and clustered sample of households in Great Britain. It is part of *Understanding Society*: The UK Household Longitudinal Study and used for methodological testing and experimentation (University of Essex, Institute for Social and Economic Research 2021). The design of the IP mirrors that of the main *Understanding Society* panel, with annual interviews of all household members aged 16+. The wave 11 interview (IP11) fielded in May to September 2018 was used as the baseline survey for Spending Study 2: a random half of respondents was invited to SS2 within the IP11 questionnaire; the other half was sent an invitation by post, a couple of weeks after completing their IP11 interview. Although the within-interview invitation increased participation in SS2 compared to the postal invitation, there was no difference in the composition of the participant samples between the two treatment groups (see Jäckle, Wenz, Burton, and Couper 2022). The invitation treatment groups are therefore combined for the purposes of the analyses presented here.

IP11 was a mixed mode survey. About two thirds of sample households were allocated to web first: all adult household members received an invitation to complete their interview online. If they did not do so after several reminders, they were followed up by a face-to-face interviewer. The remaining sample households were allocated to face-to-face interviewers, and non-respondents were given the opportunity to complete the survey online in the final stages of fieldwork. Overall, 55% of respondents completed the survey with an interviewer and 45% completed it online. For more details on IP11 fieldwork, see the Innovation Panel User Guide at https://www.understandingsociety.ac.uk/documentation/innovation-panel/user-guide.

4.2 The Lightspeed UK online access panel (AP)

Members of the Lightspeed UK online access panel were invited to complete a questionnaire which collected the same baseline information for SS2 as the IP11 questionnaire. This included socio-demographic characteristics, financial behaviours and position, mobile device access and usage, hypothetical willingness to do different types of tasks for a survey, and data security concerns. At the end of this questionnaire respondents were invited to participate in SS2 and to download the app. The Lightspeed implementation included a feedback experiment, whereby a random third of the sample were either told they would be able to see feedback about their spending within the app, were not told but able to view the feedback, or did not receive feedback. The feedback treatment had no effect on participation or reported spending, and therefore the treatment groups are combined for the purposes of the analyses presented here (Wenz, Jäckle, Burton, and Couper 2022).

4.3 Analysis sample

We restrict our analyses to participants who used the app at least once to report a purchase, which is defined as providing a non-missing purchase amount in the ‘report daily purchases’ section of the app. We use all entries participants made in the app within 31 days of first using the app.

In the IP, 2,638 sample members gave a full interview and were eligible for Spending Study 2. Of these, 446 (16.9%) used the app at least once to report a purchase, reporting a total of 12,579 purchases. In the AP, 2,878 sample members completed the baseline survey and of these, 408 (14.2%) used the app at least once, reporting a total of 11,517 purchases.

4.5 Respondent characteristics

The respondent characteristics used to examine differences in sample composition are derived from the baseline questionnaires for both samples (the IP11 questionnaire can be accessed at https://www.understandingsociety.ac.uk/documentation/innovation-panel/questionnaires. The AP baseline questionnaire is documented in Jäckle et al. 2019). The characteristics include standard socio-demographic variables, financial behaviours, and measures of digital affinity. For most variables the questions were asked in the same way in both samples. Since the rate of missing items was at most 2% in both samples, missing observations were set to the modal answer categories for the sample.

The socio-demographic characteristics included are gender (male, female), age (16-35, 36-55, 56+), highest educational qualification (degree, A/AS level [~13 years of schooling], GCSE/CE level [~11 years of schooling], no formal qualification), whether the respondent is in work (employed or self-employed in the prior week), whether living as a couple (yes, no), and the number of children aged under 16 living in the household (0, 1, 2+).

The financial behaviours include whether the respondent keeps a budget (yes, no), how often they check their bank balance (most days, at least once a week, less frequently), and whether they check their balance using an app on a mobile device (yes, no). The wording of these questions is documented in Appendix 1.

The measures of digital affinity include whether the participant uses the following devices to connect to the internet: PC or laptop (yes, no), smartphone (yes, no), tablet (yes, no); how frequently they use their smartphone (every day, less frequently); the number of different types of activities they do on their phone (summed count of 13 different activities); and self-rated smartphone skills (five-point scale coded as beginner, intermediate, advanced). The measures also include the average willingness to do different types of activities with their mobile device for a survey (four-point scales asking about willingness to do each of eight activities), and the average concerns about the security of providing data in these ways (five-point scales). In order to generate categorical indicators of willingness and data security concerns, the mean scores are rounded to the nearest integers and labelled according to the original response categories. Except for the questions about how the respondent connects to the internet, the questions were routed on using a smartphone to connect to the internet. The wording of these questions is documented in Appendix 1. For respondents who did not use a smartphone (six in the AP sample and 27 in the IP sample), the frequency of smartphone usage and number of activities are set to zero. The smartphone skills, willingness, and data security concerns are set to the lowest skill/willingness categories and to the highest concern category.

4.6 App usage behaviours

The indicators of app usage behaviour are derived from the app paradata in the same way for both samples. The indicators are coded into categories and summarise the following for each participant:

- The number of days on which the participant used the app at least once to report a direct debit or standing order, to report a purchase, or to report a 'no spend day' (coded into terciles).
- The number of direct debits or standing orders reported (coded as 0, below the median, above the median).
- The total number of daily purchases reported (coded into terciles).
- The average number of daily purchases, per day on which the participant used the app (coded as <1, 1-2, >2).
- The total number of small purchases reported, defined as costing less than £3 (coded as 0, below the median, above the median).

4.7 Measures of spending

We examine three aspects of the spending reported by participants in the app: the total value of direct debits and standing orders, the value of daily purchases by category and the total value of direct debits. The sixteen categories to record daily purchases (based on work by d'Ardenne and Blake 2012) were food and groceries; eating and drinking out; clothes and footwear; transport and car; child costs; home improvements and household goods; health expenses; socialising and hobbies; books, magazines, films and music; games and toys; haircuts, manicures and massages; holidays; gifts and donations; rent (not direct debit/standing order); bills (not direct debit/standing order); and other purchases or payments.

5. Methods

We test for differences between the two samples using $\chi 2$ tests, which account for the clustered sample design of the IP. In these analyses, we first compare the two samples to answer RQ1-RQ3. Then, we examine whether differences in the measurement of spending remain after controlling for sample composition differences and/or app usage behaviour using propensity weights (RQ4). Since some respondents of the probability sample survey were invited to the app study during a face-to-face interview instead of a web survey, we repeat all analyses using only the probability sample app participants that were invited during the web survey as a sensitivity analysis to try to account for the fact that the probability-based panel has a mixed-mode design whereas the nonprobability panel only uses the online mode of survey data collection (see Appendix Tables 2-6). Because the results with and without the app study participants invited during the face-to-face interview are broadly comparable, we only show the results for the full samples in the results section of this paper.

To examine the effectiveness of adjusting for differences in sample composition, we use propensity score weighting to match the sample composition of the nonprobability sample to that of the probability sample. We follow the approach of McCaffrey, Ridgeway and Morral (2004). To compute the weights, we combine the two samples into one dataset and estimate probit models of the probability that a participant is in the probability sample rather than the nonprobability sample. Based on these models we compute the predicted probability for each participant *i* of being in the probability sample, $p(x_i)$. The propensity score weights are then computed as $w_i = p(x_i)/[1-p(x_i)]$ for the nonprobability sample and $w_i = p(x_i)/p(x_i) = 1$ for the probability sample. The denominators of the weights adjust for the differences in characteristics between the two samples, while the numerator weights the pooled sample to match the characteristics of the probability sample: Respondents in the nonprobability sample who have characteristics that are not typical in the probability sample will have a $p(x_i)$ close to zero and therefore a weight close to zero. We estimate three separate probit models and use these to calculate three sets of weights (Appendix Table 1): weight 1 adjusts only for socio-demographic characteristics, weight 2 in addition includes financial behaviours, and weight 3 in addition includes digital affinity. The area under the curve (AUC) statistics reported in Appendix Table 1 indicate that the fit of the weighting model improves with the addition of each set of covariates: the AUC for weighting model 1 is 0.6357, for model 2 it is 0.7429, and for model 3 it is 0.8071.

6. Results

In the following, we present our results of comparing the app study samples from the probability-based mixed-mode panel and the nonprobability online panel following our research questions RQ1 to RQ4. We first focus on the unweighted estimates in Tables 1-3 to address RQ1 to RQ3.

RQ1: Do the app study participants in the probability-based panel have different characteristics than those in the nonprobability online panel?

Table 1 shows the socio-demographic characteristics that the app study participants from the two panels reported in the baseline questionnaire. When examining the unweighted app study data, we find that the nonprobability panel participants are more likely to be female, middle-aged, highly educated, and living in a household with at least one child than the probability-based panel participants, supporting H1.1. The only examined socio-demographic variables on which we did not find any significant differences between app study participants are employment status and partnership status. The average absolute difference in socio-demographic characteristics between the panels is 5.4 percentage points.

There are also significant differences in the reported financial behaviours of the app study participants (Table 2). The nonprobability panel participants are much more likely to keep a budget (75.7% versus 41.9%), check their bank balance on most days, and check their bank balance using a mobile app than the probability-based panel participants, supporting H1.4. The average absolute difference in financial behaviours between the panels is 13.7 percentage points.

Table 1: Socio-demographic characteristics

	Probability	Unweighted		Weight 1		Weight 2		Weight 3	
	sample %	Delta	P	Delta	P	Delta	P	Delta	P
Female	57.8	14.7	<0.001	0.6	0.732	4.3	0.022	1.7	0.374
Age:									
16-35	35.9	-0.3	0.014	-0.7	0.904	-3.6	0.324	1.6	0.808
36-55	41.9	6.6		1.0		4.3		-0.5	
56+	22.2	-6.3		0.4		-0.7		-1.1	
Education:									
Degree	45.3	0.3	<0.001	0.4	0.991	-0.5	0.547	-0.2	0.006
A/AS level (~13 years of schooling)	17.0	11.4		0.1		0.5		4.7	
GCSE/CSE level (~11 years of	29.8	-7.0		-0.6		1.6		-1.4	
schooling)	7.8	-4.7		0.1		-1.5		-3.1	
No formal qualification									
In work	72.4	1.4	0.503	1.2	0.556	2.2	0.282	3.2	0.112
In couple	61.9	-1.3	0.593	0.5	0.835	3.2	0.198	1.7	0.487
No. kids in household under 16:									
None	67.9	-7.9	0.025	-0.4	0.903	-2.4	0.614	0.9	0.939
One	15.9	2.7		0.9		2.0		-0.5	
Two or more	16.1	5.2		-0.5		0.4		-0.5	
Average absolute difference		5.4		0.6		2.1		1.6	

Notes: N= 408 in the nonprobability sample, N=446 in the probability sample. Weight 1 is based on socio-demographic variables only, weight 2 includes financial behaviours, and weight 3 includes digital affinity variables. Delta = percentage point difference between the nonprobability sample (not shown) and the probability sample estimate in column 1. P = p-values from χ^2 tests for differences in distributions between the two samples.

Table 2: Financial behaviours

	Probability sample	Unweighted		Weight 1		Weight 2		Weight 3	
	%	Delta	P	Delta	P	Delta	P	Delta	P
Keeps a budget	41.9	33.8	<0.001	36.0	<0.001	2.2	0.454	3.5	0.230
Checks bank balance:									
Most days	31.8	13.0	.	12.6	.	0.9	.	2.4	.
At least once a week	41.9	-1.7	.	0.1	.	0.0	.	0.5	.
Less frequently	26.2	-11.3	<0.001	-12.8	<0.001	-0.9	0.899	-2.8	0.404
Checks balance using mobile app	45.7	8.4	0.001	7.1	0.007	0.2	0.941	1.6	0.529
Average absolute difference		13.7		13.7		0.8		2.2	

Notes: N= 408 in the nonprobability sample, N=446 in the probability sample. Weight 1 is based on socio-demographic variables only, weight 2 includes financial behaviours, and weight 3 includes digital affinity variables. Delta = percentage point difference between the nonprobability sample (not shown) and the probability sample estimate in column 1. P = p-values from χ^2 tests for differences in distributions between the two samples.

Table 3 shows the digital affinity characteristics that the app study participants from the two sources reported in the baseline questionnaire. When examining the unweighted app study data, we find that the nonprobability panel participants are more likely to use desktop computers, laptops, and smartphones to connect to the internet, use a smartphone every day, and have intermediate smartphone skills. However, there are no significant differences in the use of tablets to connect to the internet and the number of activities that participants do on their smartphones. Overall, this finding partly supports H1.2. Furthermore, we find that participants in the nonprobability panel are more likely to be willing to do additional tasks on their smartphones, and be unconcerned about the security of mobile data collection than the probability-based panel participants, supporting H1.3. The average absolute difference in digital affinity between the app study samples is 7.0 percentage points.

Table 3: Digital affinity

	Probability sample %	Unweighted		Weight 1		Weight 2		Weight 3	
		Delta	P	Delta	P	Delta	P	Delta	P
Uses desktop/laptop for internet	84.5	11.8	<0.001	11.9	<0.001	11.9	<0.001	2.9	0.093
Uses Smartphone for internet	95.5	3.0	<0.001	2.4	0.001	2.8	<0.001	0.4	0.692
Uses tablet for internet	72.2	-1.4	0.546	-1.2	0.591	-2.1	0.360	0.2	0.932
Uses smartphone every day	82.3	4.2	0.018	1.2	0.542	0.9	0.631	1.8	0.354
No. activities on smartphone:									
0-9	42.6	-2.2		0.7		5.3		-0.1	
10-11	31.6	0.2		-2.5		-3.8		-3.0	
12	25.8	1.9	0.637	1.8	0.534	-1.5	0.119	-3.1	0.308
Smartphone skills:									
Beginner	10.8	-4.9		-3.3		-3.5		-2.5	
Intermediate	20.2	5.1		6.0		9.3		3.3	
Advanced	69.1	-0.2	<0.001	-2.7	0.005	-5.8	<0.001	-0.9	0.097
Willingness tasks on smartphone:	17.0								
Very	39.9	18.5		19.4		16.8		0.4	
Somewhat	43.0	8.1		7.0		6.7		0.9	
A little/not at all		-26.6	<0.001	-26.5	<0.001	-23.5	<0.001	-1.3	0.861
Data security concerns:									
Not at all	18.2	12.2		13.7		14.9		1.9	
A little	45.3	-3.4		-507		-5.5		-2.9	
Somewhat/extremely	36.5	-8.9	<0.001	-7.9	<0.001	-9.5	<0.001	1.0	0.370
Average absolute difference		7.0		7.1		7.7		1.7	

Notes: N= 408 in the nonprobability sample, N=446 in the probability sample. Weight 1 is based on socio-demographic variables only, weight 2 includes financial behaviours, and weight 3 includes digital affinity variables. Delta = percentage point difference between the nonprobability sample (not shown) and the probability sample estimate in column 1. P = p-values from χ^2 tests for differences in distributions between the two samples.

RQ2: Are there differences between the panels in how participants use the app?

Table 4 shows the app usage behaviours of the probability-based and nonprobability panel participants. When examining the unweighted app study data, we find that the nonprobability panel participants are more likely to report eight or more direct debits or standing orders and less likely to report 18-35 purchases than the probability-based panel participants. However, we did not find any significant differences in the number of days participants used the app, mean number of purchases reported per day, and the number of reported small purchases. Based on these findings, H2.1 and H2.2, that the non-probability panel participants are more likely to maximise their monetary compensation and minimiser their effort than the probability-based panel participants, are not supported. The average absolute difference in app usage characteristics between the app study samples is 3.3 percentage points.

Table 4: App use behaviours

	Probability sample %	Unweighted Delta	P	Weight 1 Delta	P	Weight 2 Delta	P	Weight 3 Delta	P
No. days used app:									
1-14	32.3	4.7		1.6		0.7		1.0	
15-25	37.0	-3.7		-1.5		0.6		1.4	
26+	30.7	-1.1	0.178	-0.1	0.783	-1.2	0.880	-2.4	0.637
No. direct debits/standing orders:									
0	26.9	-5.3		-7.7		-7.0		-7.2	
1-7	39.2	-3.0		-3.8		-2.2		-7.4	
8-21	33.9	8.3	0.004	11.5	<0.001	9.2	0.001	14.7	<0.001
No. purchases:									
1-17	31.4	4.9		4.2		4.0		0.7	
18-35	37.0	-7.6		-6.8		-2.8		1.0	
36+	31.6	-2.7	0.005	2.7	0.015	-1.3	0.203	-1.8	0.713
Mean no. purchases/day:									
<1	20.2	-1.1		-1.0		2.7		0.5	
1-2	57.6	-1.5		-1.4		-2.8		-1.1	
3+	22.2	2.6	0.466	2.3	0.533	0.1	0.378	0.7	0.895
No. small purchases (<£3):									
0	28.0	1.9		-0.1		1.1		0.8	
1-3	37.0	-1.7		-0.1		-0.7		-1.7	
4+	35.0	-0.2	0.665	0.2	0.996	-0.4	0.887	0.9	0.767
Average absolute difference		3.3		3.0		2.5		2.9	

Notes: N= 408 in the nonprobability sample, N=446 in the probability sample. Weight 1 is based on socio-demographic variables only, weight 2 includes financial behaviours, and weight 3 includes digital affinity variables. Delta = percentage point difference between the nonprobability sample (not shown) and the probability sample estimate in column 1. P = p-values from χ^2 tests for differences in distributions between the two samples.

RQ3: Are there differences between the panels in expenditure estimates, i.e., the study's main outcome of interest?

Table 5 shows the expenditures that the probability-based and nonprobability panel participants reported in the app. Not all respondents reported expenditures in all categories. Therefore, the upper panel in Table 5 examines the percentage of respondents who did not report any expenditure in a given category. The results indicate that the nonprobability panel members were more likely to report zero expenditures. For example, 11.2% of respondents in the probability panel reported zero expenditure on eating and drinking out, compared to 18.1% of the nonprobability panel members (a differences of 6.9 percentage points, $p = 0.001$). There are six other categories where the nonprobability panel members were significantly more likely than the probability panel to report zero expenditure (transport and car; home improvements and household goods; socialising and hobbies; haircuts, manicures and massages; holidays; and gifts and donations). The nonprobability panel members were however less likely than the probability panel to report zero expenditure for bills and direct debits.

The lower panel in Table 5 examines the means of reported (non-zero) expenditures. The unweighted results show significant differences in only two spending categories: nonprobability panel members reported higher mean spending on eating and drinking out (£111.6 in the probability sample, £144.8 in the nonprobability sample, a difference of £33.3, $p = 0.001$) and higher mean spending on clothes and footwear (£83.2 in the probability sample, £138.0 in the nonprobability sample, a difference of £54.7, $p < 0.001$).

That is, although there are differences in whether or not respondents report spending in a given category between the samples, the estimates of mean expenditure are similar in all but two categories. The mean total expenditure, calculated by summing up all expenditures across categories, is also similar in the two samples.

Based on these mixed findings, hypothesis H3, that estimates from the nonprobability panel would suggest lower average expenditures, has to be rejected.

Table 5: Expenditure measured in the app

	Unweighted			Weight 1		Weight 2		Weight 3	
	PrS	Delta	P	Delta	P	Delta	P	Delta	P
% of respondents reporting zero expenditure:									
Food and groceries	5.8	-0.2	0.868	-1.2	0.243	-1.1	0.305	-1.2	0.248
Eating and drinking out	11.2	**6.9**	**0.001**	**6.6**	**0.001**	**6.2**	**0.002**	3.4	0.059
Clothes and footwear	40.6	1.1	0.666	0.7	0.776	-0.8	0.752	-2.3	0.348
Transport and car	25.8	**5.1**	**0.032**	2.8	0.225	**6.1**	**0.011**	**6.0**	**0.013**
Child costs	85.7	-0.6	0.736	2.5	0.120	-0.9	0.605	0.9	0.610
Home improvements & HH goods	46.9	**5.6**	**0.033**	3.1	0.229	3.3	0.206	2.8	0.282
Health expenses	70.9	-2.5	0.290	-4.6	0.053	-3.4	0.151	**-8.6**	**0.001**
Socialising and hobbies	45.7	**17.0**	**<0.001**	**15.8**	**<0.001**	**18.2**	**<0.001**	**11.4**	**<0.001**
Books, magazines, films, and music	64.6	3.1	0.184	0.7	0.751	1.2	0.615	-2.7	0.255
Games and toys	78.5	-3.5	0.105	-2.9	0.169	**-5.3**	**0.016**	**-5.8**	**0.010**
Haircuts, manicures, massages	60.3	**11.5**	**<0.001**	**11.4**	**<0.001**	**9.0**	**<0.001**	**4.5**	**0.042**
Holidays	79.8	**8.7**	**<0.001**	**7.1**	**<0.001**	**7.9**	**<0.001**	**3.6**	**0.025**
Gifts and donations	52.9	**7.9**	**0.003**	**10.2**	**<0.001**	**9.2**	**0.001**	4.5	0.089
Rent (not direct debit/standing order)	93.3	-1.1	0.421	-2.6	0.088	**-3.3**	**0.035**	-2.0	0.167
Bills (not direct debit/standing order)	80.9	**-11.1**	**<0.001**	**-14.1**	**<0.001**	**-12.6**	**<0.001**	**-12.5**	**<0.001**
Other	42.6	-4.1	0.087	**-5.4**	**0.026**	**-8.3**	**0.001**	**-7.4**	**0.002**
Total value of direct debits	26.9	**-5.3**	**0.012**	**-7.7**	**<0.001**	**-7.0**	**0.001**	**-7.2**	**0.001**
Mean of non-zero expenditure reports:									
Total value of purchases	1041.5	-161.5	0.072	-79.1	0.375	-156.2	0.082	**-199.5**	**0.027**
Food and groceries	262.7	-27.4	0.425	-17.4	0.611	-19.1	0.578	-41.2	0.230
Eating and drinking out	111.6	**33.3**	**0.001**	**57.1**	**<0.001**	**21.6**	**0.022**	17.4	0.064
Clothes and footwear	83.2	**54.7**	**<0.001**	**32.6**	**<0.001**	**21.6**	**0.001**	4.1	0.509
Transport and car	210.2	-52.9	0.366	-49.9	0.394	-70.8	0.227	-40.1	0.492

Child costs	76.1	-6.7	0.647	-3.5	0.809	-21.4	0.146	-25.9	0.080
Home improvements & HH goods	313.0	-151.1	0.157	-105.5	0.322	-168.1	0.116	**-211.6**	**0.049**
Health expenses	56.3	8.8	0.393	**27.5**	**0.009**	8.4	0.416	-11.0	0.288
Socialising and hobbies	69.2	-13.2	0.062	-12.6	0.075	**-18.4**	**0.010**	**-17.4**	**0.015**
Books, magazines, films, and music	21.9	0.6	0.817	2.5	0.323	1.7	0.510	-2.0	0.419
Games and toys	32.8	10.6	0.067	**19.3**	**0.001**	**17.9**	**0.003**	**16.6**	**0.005**
Haircuts, manicures, massages	43.3	-2.0	0.834	-5.4	0.570	-7.0	0.461	-11.6	0.227
Holidays	340.5	41.9	0.440	10.5	0.846	15.1	0.781	-16.2	0.765
Gifts and donations	53.9	7.5	0.351	8.6	0.283	4.6	0.564	13.7	0.088
Rent (not direct debit/standing order)	293.2	40.1	0.347	29.2	0.491	51.3	0.231	8.8	0.835
Bills (not direct debit/standing order)	184.5	-42.1	0.302	-30.0	0.461	15.3	0.707	-18.4	0.651
Other	135.0	-46.4	0.188	-34.6	0.325	-30.2	0.390	-42.2	0.231
Total value of direct debits	753.1	-35.4	0.562	-54.1	0.377	-36.5	0.550	-20.5	0.737

Notes: N= 408 in the nonprobability sample, N=446 in the probability sample. Weight 1 is based on socio-demographic variables only, weight 2 includes financial behaviours, and weight 3 includes digital affinity variables. Delta = percentage point difference between the nonprobability sample (not shown) and the probability sample estimate in column 1. In the upper panel, percent of respondents reporting £0 expenditure in the given category, P = p-values from χ^2 tests for differences in distributions between the two samples. In the lower panel, mean of non-zero expenditure reports, P = p-value from tests of differences in means between the two samples.

RQ4: Do the differences between panels in expenditure estimates remain after weighting?

The results in Table 5 suggest that none of the weights reduce the differences between panels in expenditure estimates. Focusing on total mean expenditure, the unweighted estimates are not significantly different between the two samples. Neither the first weight (based on socio-demographic characteristics), nor the second weight (which in addition includes financial behaviours) change this conclusion. However, using the third weight (which in addition includes digital affinity), suggests a significantly lower mean total expenditure in the nonprobability panel compared to the probability panel (-£199.5, $p = 0.027$).

Examining individual expenditure categories, the results are mixed. The differences between samples in mean expenditure on eating and drinking out, and clothing and footwear, become insignificant when weight 3 is used. On the other hand, significant differences in mean expenditure emerge when weights are applied for home improvements and household goods, health expenses, socialising and hobbies, and games and toys.

These results do not support the hypotheses that differences between probability and nonprobability samples will decrease when weighting for socio-demographic characteristics (H4.1), decrease further when adding financial behaviours (H4.2), and decrease further still when adding digital affinity measures related to the app use task (H4.3).

7. Conclusions

In this study, we examined differences between app study data gathered based on a probability-based mixed-mode panel and a nonprobability online panel. We found that app study participants from the two panels differed on socio-demographics, financial behaviour and digital affinity as well as their spending reported in the app. These findings show that different people participated in the two app study samples and that the data from the two app study samples led to different conclusions regarding key substantive app study outcomes. Weighting did not reduce differences in mean expenditure between samples, even when the weighting scheme included extensive participant information from the baseline questionnaire.

In line with most of the research on nonprobability and probability-based panel data, our results indicate that the differences in the nonprobability and probability-based panel recruitment processes lead to differences in the data gathered (e.g., socio-demographic characteristics, financial behaviour, digital skills and behaviour, and substantive outcomes of interest) from these data sources. In addition, in line with most research on nonprobability and probability-based panel data, our findings show that these differences in the data are difficult to eliminate by weighting. While most existing research only examined survey data, our findings indicate that the differences can also be found in app study data gathered from nonprobability and probability-based panel respondents.

The only data quality aspect for which we did not find evidence of differences between the nonprobability and probability-based panel was behaviour in using the spending app. This finding is contrary to the argument that nonprobability panel participants try to maximize their monetary incentive at the expense of data quality.

However, this finding is in line with some of the scarce existing literature on response behaviour in surveys which is inconclusive regarding the question of whether nonprobability panel participants answer questions less conscientiously than probability-based panel respondents.

Since our study is the first to compare app study data based on nonprobability and probability-based panels, more research is needed to verify our findings and to further explore which features of nonprobability and probability-based panels lead to these differences in the gathered data. The extent to which our findings are generalisable beyond spending apps and beyond the specific panel survey studies (*Understanding Society* Innovation Panel and Lightspeed UK) also remains to be explored in future research. In addition, future research should examine which sample source for gathering app study data leads to more accurate estimates relative to benchmark data for the intended target population. Based on our study, we conclude that it remains critical to be cautious about app studies conducted in a nonprobability panel, since results may not translate to app studies conducted in a probability-based panel setting.

References

Ansolabehere, S., and Rivers, D. (2013), "Cooperative Survey Research," *Annual Review of Political Science*, 16, 307–329.

Ansolabehere, S., and Schaffner, B. F. (2014), "Does Survey Mode Still Matter? Findings from a 2010 Multi-Mode Comparison," *Political Analysis*, 22(3), 285–303.

Berrens, R. P., Bohara, A. K., Jenkins-Smith, H., Silva, C., and Weimer, D. L. (2003), "The Advent of Internet Surveys for Political Research: A Comparison of Telephone and Internet Samples," *Political Analysis*, 11(1), 1–22.

Brüggen, E., van den Brakel, J., and Krosnick, J. (2016), *Establishing the Accuracy of Online Panels for Survey Research. Statistics Netherlands Discussion Paper 2016-04*, The Hague: Statistics Netherlands. Available at https://www.cbs.nl/en-gb/background/2016/15/establishing-the-accuracy-of-online-panels-for-survey-research.

Callegaro, M., Baker, R., Bethlehem, J., Göritz, A. S., Krosnick, J. A., and Lavrakas, P. J. (2014), "Online Panel Research: History, Concepts, Applications and a Look at the Future," In Callegaro, M. Baker, J., Göritz, A., Krosnick, J. A., and Lavrakas, P. J. (Eds.), *Online Panel Research: A Data Quality Perspective* (pp. 1–22), Chichester: John Wiley and Sons.

Chan, P., and Ambrose, D. (2011), "Canadian Online Panels: Similar or Different," *Vue,* 16–20.

Chang, L., and Krosnick, J. A. (2009), "National Surveys via RDD Telephone Interviewing versus the Internet: Comparing Sample Representativeness and Response Quality," *Public Opinion Quarterly*, 73(4), 641–678.

Cornesse, C., and Blom, A. G. (2020), "Response Quality in Nonprobability and Probability-Based Online Panels," *Sociological Methods and Research*, 1–30.

Cornesse, C., Blom, A. G., Dutwin, D., Krosnick, J. A., de Leeuw, E. D., Legleye, S., Pasek, J., Pennay, D., Phillips, B., Sakshaug, J. W., Struminskaya, B., and Wenz, A. (2020), "A Review of Conceptual Approaches and Empirical Evidence on Probability and Nonprobability Sample Survey Research," *Journal of Survey Statistics and Methodology*, 8(1), 4-36.

D'Ardenne, J., and Blake, M. (2012), *Developing Expenditure Questions: Findings from Focus Groups. IFS Working Paper W12/18*, London: Institute for Fiscal Studies. Available at https://ifs.org.uk/publications/developing-expenditure-questions-findings-focus-groups.

Dassonneville, R., Blais, A., Hooghe, M., and Deschouwer, K. (2020), "The Effects of Survey Mode and Sampling in Belgian Election Studies: A Comparison of a National Probability Face-to-Face Survey and a Nonprobability Internet Survey," *Acta Politica*, 55, 175–198.

Duffy, B., Smith, K., Terhanian, G., and Bremer, J. (2005), "Comparing Data from Online and Face-to-Face Surveys," *International Journal of Market Research*, *47*(6), 615–639.

Dutwin, D., and Buskirk, T. D. (2017), "Apples to Oranges or Gala versus Golden Delicious? Comparing Data Quality of Nonprobability Internet Samples to Low Response Rate Probability Samples," *Public Opinion Quarterly*, 81(S1), 213–239.

Erens, B., Burkill, S., Couper, M. P., Conrad, F., Clifton, S., Tanton, C., Phelps, A., Datta, J., Mercer, C. H., Sonnenberg P., Prah P., Mitchell K. R., Wellings K., Johnson A. M., and Copas, A. J. (2014), "Nonprobability Web Surveys to

Measure Sexual Behaviors and Attitudes in the General Population: A Comparison with a Probability Sample Interview Survey," *Journal of Medical Internet Research*, 16(12): e276.

Gelman, A., Goel, S., Rothschild, D., and Wang, W. (2016), "High-Frequency Polling with Non-Representative Data." In Schill, D., Kirk, R., and Jasperson, A. E. (Eds.) *Political Communication in Real Time: Theoretical and Applied Research Approaches* (pp. 117–133), New York: Routledge.

Gittelman, S. H., Thomas, R. K., Lavrakas, P. J., and Lange, V. (2015), "Quota Controls in Survey Research: A Test of Accuracy and Intersource Reliability in Online Samples," *Journal of Advertising Research*, 55, 368–379.

Greszki, R., Meyer, M., and Schoen, H. (2014), "The Impact of Speeding on Data Quality in Nonprobability and Freshly Recruited Probability-Based Online Panels," In Callegaro, M. Baker, J., Göritz, A., Krosnick, J. A., and Lavrakas, P. J. (Eds.), *Online Panel Research: A Data Quality Perspective* (pp. 238–262), Chichester: John Wiley and Sons.

Groves, R. M., and Couper, M. P. (1998). *Nonresponse in Household Interview Surveys*, New York: John Wiley and Sons.

Hillygus, D. S., Jackson, N., and Young, M. (2014), "Professional Respondents in Nonprobability Online Panels," In Callegaro, M. Baker, J., Göritz, A., Krosnick, J. A., and Lavrakas, P. J. (Eds.), *Online Panel Research: A Data Quality Perspective* (pp. 219–237), Chichester: John Wiley and Sons.

Jäckle, A., Burton, J., Wenz, A., and Read, B. (2018), "Understanding Society The UK Household Longitudinal Study: Spending Study 1 User Guide," Colchester: University of Essex.

Jäckle, A., Burton, J., Wenz, A., Read, B., Hanson, T., and Xu, D. (2019), *Understanding Society: The UK Household Longitudinal Study. Spending Study 2, User Guide*, Colchester: Institute for Social and Economic Research, University of Essex. Available from the UK Data Service at https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8909.

Jäckle, A., Wenz, A., Burton, J., and Couper, M. P. (2022), "Increasing Participation in a Mobile App Study: The Effects of a Sequential Mixed-Mode Design and In-Interview Invitation", *Journal of Survey Statistics and Methodology*, 10(4), 898–922.

Kennedy, C., Mercer, A., Keeter, S., Hatley, N., McGeeney, K., and Gimenez, A. (2016), *Evaluating Online Nonprobability Surveys: Vendor Choice Matters; Widespread Errors Found for Estimates Based on Blacks and Hispanics*, Washington DC: Pew Research Center. Available at https://www.pewresearch.org/methods/2016/05/02/evaluating-online-nonprobability-surveys/.

Keusch, F., Batinic, B., and Mayerhofer, W. (2014), "Motives for Joining Nonprobability Online Panels and Their Association with Survey Participation Behavior," In Callegaro, M. Baker, J., Göritz, A., Krosnick, J. A., and Lavrakas, P. J. (Eds.), *Online Panel Research: A Data Quality Perspective* (pp. 171–191), Chichester: John Wiley and Sons.

Lee, S. (2006), "Propensity Score Adjustment as a Weighting Scheme for Volunteer Panel Web Surveys," *Journal of Official Statistics*, 22(2), 329–349.

Legleye, S., Charrance, G., Razafindratsima, N., Bajos, N., Bohet, A., Moreau, C., and the FECOND Research Team (2018), "The Use of a Nonprobability

Internet Panel to Monitor Sexual and Reproductive Health in the General Population," *Sociological Methods and Research*, 47(2), 314–348.

Loosveldt, G., and Sonck, N. (2008), "An Evaluation of the Weighting Procedures for an Online Access Panel Survey," *Survey Research Methods*, 2(2), 93–105.

MacInnis, B., Krosnick, J. A., Ho, A. S., and Cho, M.-J. (2018), "The Accuracy of Measurements with Probability and Nonprobability Survey Samples: Replication and Extension", *Public Opinion Quarterly,* 82(4), 707–744.

Malhotra, N., and Krosnick, J. A. (2007), "The Effect of Survey Mode and Sampling on Inferences about Political Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys with Nonprobability Samples," *Political Analysis*, 15(3), 286–323.

McCaffrey, D.F., Ridgeway, G., and Morral, A.R. (2004), "Propensity Score Estimation with Boosted Regression for Evaluating Causal Effects in Observational Studies," *Psychological Methods*, 9(4), 403–425.

Mercer, A. W., Kreuter, F., Keeter, S., and Stuart, E. A. (2017), "Theory and Practice in Nonprobability Surveys: Parallels between Causal Inference and Survey Inference," *Public Opinion Quarterly,* 81(S1), 250–271.

Pasek, J. (2016), "When Will Nonprobability Surveys Mirror Probability Surveys? Considering Types of Inference and Weighting Strategies as Criteria for Correspondence," *International Journal of Public Opinion Research,* 28(2), 269–291.

Pennay, D. W., Neiger, D., Lavrakas, P. J., and Borg, K. (2018), *The Online Panels Benchmarking Study: A Total Survey Error Comparison of Findings from Probability-Based Surveys and Non-Probability Online Panel Surveys in Australia. CSRM & SRC Methods Paper No. 2/2018*, Canberra: Australian

National University, Centre for Social Research and Methods. Available at https://csrm.cass.anu.edu.au/research/publications/online-panels-benchmarking-study-total-survey-error-comparison-findings.

Scherpenzeel, A. C., and Bethlehem, J. G. (2011), “How Representative Are Online Panels? Problems of Coverage and Selection and Possible Solutions,” In Das, M., Ester, P., and Kaczmirek, L. (Eds.) *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies* (pp. 105–132), New York: Routledge.

Schonlau, M., van Soest, A., and Kapteyn, A. (2007), “Are ‘Webographic’ or Attitudinal Questions Useful for Adjusting Estimates from Web Surveys Using Propensity Scoring?,” *Survey Research Methods*, 1(3), 155–163.

Schonlau, M., van Soest, A., Kapteyn, A., and Couper, M. P. (2009), “Selection Bias in Web Surveys and the Use of Propensity Scores,” *Sociological Methods and Research,* 37(3), 291–318.

Schonlau, M., Zapert, K., Simon, L. P., Sanstad, K. H., Marcus, S. M., Adams, J., Spranca, M., Kan, H., Turner, R., and Berry, S. H. (2004), “A Comparison between Responses from a Propensity-Weighted Web Survey and an Identical RDD Survey,” *Social Science Computer Review*, 22(1), 128–138.

Smyk, M., Tyrowicz, J., and van der Velde, L. (2021), “A Cautionary Note on the Reliability of the Online Survey Data: The Case of Wage Indicator,” *Sociological Methods and Research*, 50(1), 429–464.

Sohlberg, J., Gilljam, M., and Martinsson, J. (2017), “Determinants of Polling Accuracy: The Effect of Opt-In Internet Surveys,” *Journal of Elections, Public Opinion and Parties,* 27(4), 433–447.

Sparrow, N. (2006), “Developing Reliable Online Polls,” *International Journal of Market Research,* 48(6), 659–680.

Steinmetz, S., Bianchi, A., Tijdens, K., and Biffignandi, S. (2014), “Improving Web Survey Quality: Potentials and Constraints of Propensity Score Adjustments,” In Callegaro, M. Baker, J., Göritz, A., Krosnick, J. A., and Lavrakas, P. J. (Eds.), *Online Panel Research: A Data Quality Perspective* (pp. 273–298), Chichester: John Wiley and Sons.

Sturgis, P., Kuha, J., Baker, N., Callegaro, M., Fisher, S., Green, J., Jennings, W., Lauderdale, B. E., and Smith, P. (2018), “An Assessment of the Causes of the Errors in the 2015 UK General Election Opinion Polls,” *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 181(3), 757–781.

Suffield, M., Hasbrouck, H., Coulter, A., Jäckle, A., Burton, J., Crossley, T. F., Couper, M. P., and Lessof, C. (2018), *Understanding How People Think about Their Daily Spending. Understanding Society Working Paper 2018-02*, Colchester: University of Essex.

Szolnoki, G., and Hoffmann, D. (2013), “Online, Face-to-Face and Telephone Surveys— Comparing Different Sampling Methods in Wine Consumer Research,” *Wine Economics and Policy*, 2(2), 57–66.

University of Essex, Institute for Social and Economic Research. (2021). *Understanding Society: Innovation Panel, Waves 1-13, 2008-2020*. [data collection]. *11th Edition.* UK Data Service. SN: 6849, DOI: 10.5255/UKDA-SN-6849-14.

University of Essex, Institute for Social and Economic Research (2022), *Understanding Society: Spending Study 2, 2018-2019.* [data collection]. UK Data Service. SN: 8909, DOI: 10.5255/UKDA-SN-8909-1.

Wang, W., Rothschild, D., Goel, S., and Gelman, A. (2015), “Forecasting Elections with Non-Representative Polls,” *International Journal of Forecasting,* 31(3), 980–991.

Wenz, A., Jäckle, A., Burton, J., and Couper, M. P. (2022), “The Effects of Personalized Feedback on Participation and Reporting in Mobile App Data Collection,” *Social Science Computer Review*, 40(1), 165–178.

Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A., and Wang, R. (2011), “Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples,” *Public Opinion Quarterly, 75*(4), 709–747.

Zindel, Z. (2022), “Social Media Recruitment in Online Survey Research: A Systematic Literature Review,” *methods, data, analyses*, *42*, 1–42.

Appendix 1: Question wording

The questions in the Innovation Panel (IP) and the access panel (AP) were the same unless noted otherwise below.

Digital affinity

Access to mobile technology

Which of the following devices do you use to connect to the Internet? (Yes, No)

Desktop or laptop computer

Smartphone

Tablet

Other

Frequency of smartphone use [Ask if Smartphone = Yes]

How often do you use a **smartphone** for activities other than phone calls or text messaging?

Every day

Several times a week

Several times a month

Once a month or less

Activities on smartphone [Ask if Smartphone = Yes] – IP VERSION

Do you use your **smartphone** for the following activities? (Yes, No)

Browsing websites

Email

Taking photos

Looking at content on social media websites/apps (e.g., looking at text, images, videos on Facebook, Twitter, Instagram)

Posting content to social media websites/apps (e.g., posting text, images, videos on Facebook, Twitter, Instagram)

Making purchases (e.g., booking train tickets, buying clothes, ordering food)

Online banking (e.g., checking account balance, transferring money)

Installing new apps (e.g., from iTunes, Google Play Store)

Using GPS/location-aware apps (e.g., Google Maps, Foursquare, Yelp)

Connecting to other electronic devices via Bluetooth (e.g., smartwatches, bathroom scales)

Playing games

Streaming videos or music

Other

Activities on smartphone [Ask if Smartphone = Yes] – AP VERSION

Do you use your **smartphone** for the following activities? (Yes, No, Don't know)

Browsing websites

Email

Taking photos

Looking at content on social media websites/apps

Posting content to social media websites/apps

Making purchases

Online banking

Installing new apps

Using GPS/location-aware apps

Connecting to other electronic devices via Bluetooth

Playing games

Streaming videos or music

Other

Self-reported smartphone skills [Ask if Smartphone = Yes]

Generally, how would you rate your skills of using a **smartphone** on a scale from 1 = Beginner to 5 = Advanced?

1 Beginner

2

3

4

5 Advanced

Willingness to participate [Ask if Smartphone = Yes] – IP VERSION

How willing would you be to carry out the following tasks on your **smartphone** for a survey?

(Very willing, Somewhat willing, A little willing, Not at all willing)

Complete an online questionnaire on your mobile phone

Download a survey app to complete an online questionnaire

Download an app which collects anonymous data about how you use your smartphone

Answer a couple of questions sent via text messaging

Use the camera to take photos or scan barcodes

Allow built-in features of your smartphone to measure the frequency and speed at which you walk, run or cycle

Share the GPS position of your smartphone

Connect your smartphone via Bluetooth to other electronic devices (e.g., wearables such as Fitbit)

Willingness to participate [Ask if Smartphone = Yes] – AP VERSION

How willing would you be to carry out the following tasks on your **smartphone** for a survey?

(Very willing, Somewhat willing, A little willing, Not at all willing)

Complete an online questionnaire

Download a survey app to complete an online questionnaire

Download an app which collects anonymous data about how you use your smartphone

Answer a couple of questions sent via text messaging

Use the camera to take photos or scan barcodes

Allow built-in features to measure the frequency and speed at which you walk, run or cycle

Share the GPS position of your smartphone

Connect via Bluetooth to other electronic devices

Privacy/security concerns [Ask if Smartphone = Yes or Tablet = Yes] – IP VERSION

In general, how concerned would you be about the security of providing information in the following ways?

(Not at all concerned, A little concerned, Somewhat concerned, Very concerned, Extremely concerned)

Complete an online questionnaire in your mobile browser

Download a survey app to complete an online questionnaire

Download an app which collects anonymous data about how you use your <smartphone/tablet/smartphone or tablet>

Answer a couple of questions sent via text messaging [Ask if Smartphone = Yes]

Use the camera of your <smartphone/tablet/smartphone or tablet> to take photos or scan barcodes

Allow built-in features of your smartphone to measure the frequency and speed at which you walk, run or cycle [Ask if Smartphone = Yes]

Share the GPS position of your smartphone [Ask if Smartphone = Yes]

Connect your <smartphone/tablet/smartphone or tablet> via Bluetooth to other electronic devices (e.g., wearables such as Fitbit)

Privacy/security concerns [Ask if Smartphone = Yes or Tablet = Yes] – AP VERSION

In general, how concerned would you be about the security of providing information in the following ways?

(Not at all concerned, A little concerned, Somewhat concerned, Very concerned, Extremely concerned)

Complete an online questionnaire in your mobile browser

Download a survey app to complete an online questionnaire

Download an app which collects anonymous data about how you use your <smartphone/tablet/smartphone or tablet>

Answer a couple of questions sent via text messaging [Ask if Smartphone = Yes]

Use the camera of your <smartphone/tablet/smartphone or tablet> to take photos or scan barcodes

Allow built-in features of your smartphone to measure the frequency and speed at which you walk, run or cycle [Ask if Smartphone = Yes]

Share the GPS position of your smartphone [Ask if Smartphone = Yes]

Connect your <smartphone/tablet/smartphone or tablet> via Bluetooth to other electronic devices

Financial behaviours

Method of budgeting – IP VERSION

Now, thinking about different ways that people have of managing their finances, how do you keep your budget? Please select all that apply.

On a paper

On a computer document or spreadsheet (e.g., using Microsoft Excel)

Using personal budget software installed on a computer or laptop (e.g., Microsoft Money)

Using an online personal budget programme (e.g., Money Dashboard)

Using a personal budget app on a mobile device (e.g., You Need A Budget-YNAB)

I don't keep a budget

Frequency of checking bank balance

How often do you check your bank balance?

Most days

At least once a week

A couple of times a month

At least once a month

Less than once a month

Never

Method of checking bank balance

How do you check your bank balance? Please select all that apply.

At a cashpoint/ATM

On-line

By telephone

Using an app on a mobile device

Text messages/alerts from the bank

Paper statement

None of the above

Appendix 2: Weighting

Appendix Table 1: Weighting models (probability of being in the probability sample)

Probit models	Model 1		Model 2		Model 3	
	Coef	S.E.	Coef	S.E.	Coef	S.E.
Female	-0.391***	0.094	-0.442***	0.100	-0.453***	0.106
(Age 16-35)						
Age 36-55	-0.156	0.100	-0.228*	0.106	-0.218	0.118
Age 56+	-0.012	0.138	-0.095	0.147	-0.049	0.171
(Degree)						
A/AS level	-0.342**	0.114	-0.359**	0.120	-0.442***	0.126
GCSE/CSE level	0.138	0.110	0.092	0.116	0.025	0.125
No formal qualification	0.524*	0.213	0.584*	0.228	0.431	0.248
In work	0.059	0.109	0.127	0.115	0.167	0.124
(Number of kids: 0)						
Number of kids: 1	-0.116	0.125	-0.076	0.130	-0.101	0.140
Number of kids: 2+	-0.207	0.124	-0.209	0.129	-0.230	0.139
In couple	0.080	0.096	0.145	0.102	0.163	0.109
Keeps a budget			-0.915***	0.096	-0.851***	0.102
(Checks balance most days)						
Checks at least once a week			0.120	0.105	0.102	0.112
Checks less frequently			0.361**	0.134	0.373*	0.145
Checks balance w/ mobile app			-0.112	0.098	-0.102	0.107
Has desktop computer					-0.838***	0.190
Has smartphone					-0.449	0.378
Has tablet					0.116	0.111
Uses SP less than every day					-0.042	0.169
(No. activities on SP: 0-9)						
No. activities on SP: 10-11					0.378**	0.133
No. activities on SP: 12					0.365*	0.149
(SP skills: beginner)						
SP skills: intermediate					-0.130	0.248
SP skills: advanced					0.205	0.248
(Very willing to do tasks on SP)						
Somewhat willing					0.331*	0.135
A little/not at all willing					1.203***	0.180
(Not at all concerned)						
A little concerned					0.223	0.136
Somewhat/extremely concerned					-0.146	0.166
Constant	0.363*	0.160	-0.074	0.196	0.216	0.482
N	854		854		854	
AUC	0.6357		0.7429		0.8071	

Notes: SP=smartphone.

Appendix 3: Comparison using only the IP web respondents

Appendix Table 2: Socio-demographic characteristics

	Probability sample	Unweighted		Weight 1		Weight 2		Weight 3	
	%	Delta	P	Delta	P	Delta	P	Delta	P
Female	56.9	15.6	<0.001	1.6	0.515	5.2	0.031	2.6	0.283
Age:									
16-35	35.6	0.0	.	-0.4	.	-3.3	.	1.9	.
36-55	42.3	6.2	.	-0.7	.	3.9	.	-0.9	.
56+	22.1	-6.2	0.064	1.1	0.918	-0.6	0.516	-1.1	0.824
Education:									
Degree	47.4	-1.8	.	-1.7	.	-2.6	.	-2.4	.
A/AS level (~13 years of schooling)	14.2	14.2	.	2.9	.	3.3	.	7.5	.
GCSE/CSE level (~11 years of schooling)	28.9	-6.1	.	0.3	.	2.5	.	-0.4	.
No formal qualification	9.5	-6.3	<0.001	-1.5	0.522	-3.2	0.125	-4.7	0.002
In work	73.9	-0.1	0.959	-0.3	0.910	0.7	0.803	1.7	0.525
In couple	64.4	-3.9	0.206	-2.0	0.504	0.6	0.835	-0.8	0.785
No. kids in household under 16:									
None	71.9	-11.9	.	-4.4	.	-6.4	.	-3.1	.
One	10.7	8.0	.	6.1	.	7.2	.	4.9	.
Two or more	17.4	3.9	0.010	-1.7	0.115	-0.8	0.059	-1.7	0.225
Average absolute difference		6.5		1.9		3.1		2.6	

Notes: N= 408 in the nonprobability sample, N=253 in the probability sample. Weight 1 is based on socio-demographic variables only, weight 2 includes financial behaviours, and weight 3 includes digital affinity variables. Delta = percentage point difference between the nonprobability sample (not shown) and the probability sample estimate in column 1. P = p-values from χ^2 tests for differences in distributions between the two samples.

Appendix Table 3: Financial behaviours

	Probability sample %	Unweighted		Weight 1		Weight 2		Weight 3	
		Delta	P	Delta	P	Delta	P	Delta	P
Keeps a budget	45.8	29.9	<0.001	32.1	<0.001	-1.7	0.642	-0.4	0.920
Checks bank balance:									
Most days	33.6	11.3	.	10.9	.	-0.8	.	0.6	.
At least once a week	39.1	1.1	.	2.9	.	2.8	.	3.3	.
Less frequently	27.3	-12.3	<0.001	-13.8	<0.001	-1.9	0.696	-3.9	0.424
Checks balance using mobile app	52.6	1.6	0.604	0.3	0.930	-6.6	0.034	-5.2	0.094
Average absolute difference		11.2		12.0		2.8		2.7	

Notes: N= 408 in the nonprobability sample, N=253 in the probability sample. Weight 1 is based on socio-demographic variables only, weight 2 includes financial behaviours, and weight 3 includes digital affinity variables. Delta = percentage point difference between the nonprobability sample (not shown) and the probability sample estimate in column 1. P = p-values from χ^2 tests for differences in distributions between the two samples.

Appendix Table 4: Digital affinity

	Probability sample %	Unweighted		Weight 1		Weight 2		Weight 3	
		Delta	P	Delta	P	Delta	P	Delta	P
Uses desktop/laptop for internet	83.4	12.9	<0.001	13.0	<0.001	13.0	<0.001	4.0	0.097
Uses Smartphone for internet	95.7	2.9	<0.001	2.3	0.017	2.6	0.003	0.2	0.857
Uses tablet for internet	72.3	-1.5	0.616	-1.3	0.652	-2.2	0.461	0.1	0.985
Uses smartphone every day	84.6	1.9	0.383	-1.1	0.640	-1.4	0.572	-0.5	0.819
No. activities on smartphone:									
0-9	43.5	-3.0	.	-0.2	.	4.4	.	-1.0	.
10-11	27.3	4.6	.	1.8	.	0.6	.	1.3	.
12	29.2	-1.6	0.281	-1.6	0.759	-5.0	0.171	-0.3	0.885
Smartphone skills:									
Beginner	10.3	-4.4	.	-2.8	.	-3.0	.	-2.0	.
Intermediate	19.8	5.5	.	6.4	.	9.7	.	3.8	.
Advanced	70.0	-1.1	0.007	-3.6	0.030	-6.7	0.001	-1.8	0.232
Willingness tasks on smartphone:									
Very	18.2	17.4	.	18.3	.	15.7	.	-0.8	.
Somewhat willing	42.3	5.7	.	4.6	.	4.3	.	-1.5	.
A little/not at all willing	39.5	-23.1	<0.001	-22.9	<0.001	-20.0	<0.001	2.2	0.755
Data security concerns:									
Not at all	18.2	12.2	.	13.7	.	14.9	.	1.9	.
A little concerned	45.5	-3.5	.	-5.9	.	-5.6	.	-3.1	.
Somewhat/extremely concerned	36.4	-8.7	<0.001	-7.8	<0.001	-9.3	<0.001	1.2	0.540

Average absolute difference	6.9	6.7	7.4	1.6

Notes: N= 408 in the nonprobability sample, N=253 in the probability sample. Weight 1 is based on socio-demographic variables only, weight 2 includes financial behaviours, and weight 3 includes digital affinity variables. Delta = percentage point difference between the nonprobability sample (not shown) and the probability sample estimate in column 1. P = p-values from χ^2 tests for differences in distributions between the two samples.

Appendix Table 5: App use behaviours

	Probability sample	Unweighted		Weight 1		Weight 2		Weight 3	
	%	Delta	P	Delta	P	Delta	P	Delta	P
No days used app:									
1-14	32.4	4.6	.	1.5	.	0.5	.	0.8	.
15-25	34.0	-0.7	.	1.5	.	3.6	.	4.4	.
26+	33.6	-3.9	0.295	-2.9	0.630	-4.1	0.366	-5.2	0.208
No. direct debits/standing orders:									
0	27.7	-6.1	.	-8.4	.	-7.8	.	-8.0	.
1-7	42.3	-6.0	.	-6.9	.	-5.2	.	-10.5	.
8-21	30.0	12.1	0.002	15.3	<0.001	13.0	0.001	18.5	<0.001
No. purchases:									
1-17	32.0	4.3	.	3.5	.	3.4	.	0.1	.
18-35	37.2	-7.7	.	-7.0	.	-3.0	.	0.9	.
36+	30.8	3.5	0.044	3.4	0.080	-0.5	0.503	-1.0	0.941
Mean no. purchases/day:									
<1	22.1	-3.0	.	-2.9	.	0.8	.	-1.5	.
1-2	55.3	0.8	.	0.9	.	-0.5	.	1.1	.
>2	22.5	2.2	0.489	2.0	0.524	-0.3	0.964	0.4	0.866
No. small purchases (<£3):									
0	28.5	1.4	.	-0.5	.	0.6	.	0.4	.
1-3	38.7	-3.4	.	-1.8	.	-2.4	.	-3.5	.
4+	32.8	2.0	0.472	2.4	0.673	1.8	0.669	3.1	0.412
Average absolute difference		4.1		4.1		3.2		4.0	

Notes: N= 408 in the nonprobability sample, N=253 in the probability sample. Weight 1 is based on socio-demographic variables only, weight 2 includes financial behaviours, and weight 3 includes digital affinity variables. Delta = percentage point difference between the nonprobability sample (not shown) and the probability sample estimate in column 1. P = p-values from χ^2 tests for differences in distributions between the two samples.

Appendix Table 6: Expenditure measured in the app

	Unweighted			Weight 1		Weight 2		Weight 3	
	PrS	Delta	P	Delta	P	Delta	P	Delta	P
% of respondents reporting zero expenditure:									
Food and groceries	6.3	-0.7	0.638	-1.7	0.194	-1.6	0.237	-1.7	0.197
Eating and drinking out	12.3	**5.9**	**0.021**	**5.6**	**0.027**	**5.2**	**0.039**	2.4	0.303
Clothes and footwear	38.7	2.9	0.363	2.6	0.426	1.1	0.740	-0.5	0.879
Transport and car	23.3	**7.6**	**0.011**	5.3	0.067	**8.6**	**0.004**	**8.5**	**0.005**
Child costs	86.6	-1.5	0.500	1.6	0.427	-1.8	0.416	0.0	0.985
Home improvements & HH goods	47.4	5.0	0.135	2.6	0.446	2.7	0.417	2.2	0.506
Health expenses	72.3	-3.9	0.208	-6.1	0.057	-4.9	0.125	**-10.0**	**0.003**
Socialising and hobbies	49.0	**13.7**	**<0.001**	**12.5**	**<0.001**	**14.9**	**<0.001**	**8.1**	**0.014**
Books, magazines, films, and music	66.8	0.8	0.765	-1.5	0.608	-1.0	0.716	-5.0	0.096
Games and toys	82.2	**-7.2**	**0.007**	**-6.7**	**0.011**	**-9.1**	**0.001**	**-9.5**	**0.001**
Haircuts, manicures, massages	56.1	**15.7**	**<0.001**	**15.6**	**<0.001**	**13.2**	**<0.001**	**8.7**	**0.020**
Holidays	77.5	**11.0**	**<0.001**	**9.5**	**<0.001**	**10.2**	**<0.001**	**6.0**	**0.016**
Gifts and donations	50.2	**10.6**	**0.001**	**12.9**	**<0.001**	**11.9**	**<0.001**	**7.2**	**0.026**
Rent (not direct debit/standing order)	93.3	-1.1	0.505	-2.6	0.159	-3.3	0.082	-2.0	0.254
Bills (not direct debit/standing order)	80.2	**-10.4**	**<0.001**	**-13.4**	**<0.001**	**-11.9**	**<0.001**	**-11.8**	**<0.001**
Other	39.9	-1.4	0.623	-2.7	0.356	-5.6	0.052	-4.7	0.105
Total value of direct debits	27.7	-6.1	0.027	**-8.4**	**0.002**	**-7.8**	**0.004**	**-8.0**	**0.003**
Mean of non-zero expenditure reports:									

Total value of purchases	1093.9	-213.8	0.114	-131.5	0.329	-208.6	0.123	-251.9	0.063
Food and groceries	232.2	3.0	0.917	13.0	0.657	11.3	0.698	-10.8	0.712
Eating and drinking out	106.0	**38.8**	**<0.001**	**62.6**	**<0.001**	**27.1**	**0.002**	**22.9**	**0.009**
Clothes and footwear	83.3	**54.6**	**<0.001**	**32.5**	**<0.001**	**21.5**	**0.006**	4.0	0.599
Transport and car	254.9	-97.7	0.299	-94.6	0.314	-115.5	0.220	-84.8	0.366
Child costs	57.2	12.2	0.339	15.4	0.231	-2.5	0.842	-7.1	0.578
Home improvements & HH goods	376.0	-214.1	0.228	-168.5	0.342	-231.0	0.194	-274.5	0.124
Health expenses	43.8	21.4	0.054	**40.0**	**0.001**	20.9	0.058	1.5	0.890
Socialising and hobbies	69.0	-13.0	0.229	-12.4	0.250	-18.2	0.094	-17.2	0.114
Books, magazines, films, and music	22.5	0.0	0.990	1.9	0.565	1.0	0.750	-2.6	0.414
Games and toys	27.1	**16.3**	**0.026**	**25.0**	**0.001**	**23.6**	**0.002**	**22.3**	**0.003**
Haircuts, manicures, massages	35.4	5.9	0.177	2.5	0.571	0.8	0.845	-3.7	0.395
Holidays	317.5	64.9	0.345	33.5	0.624	38.1	0.578	6.9	0.920
Gifts and donations	64.8	-3.4	0.784	-2.3	0.856	-6.3	0.617	2.9	0.818
Rent (not direct debit/standing order)	356.2	-23.0	0.691	-33.8	0.559	-11.8	0.838	-54.3	0.353
Bills (not direct debit/standing order)	191.0	-48.5	0.280	-36.4	0.416	8.8	0.843	-24.8	0.579
Other	151.2	-62.6	0.282	-50.8	0.382	-46.5	0.424	-58.4	0.315
Total value of direct debits	777.5	-59.8	0.529	-78.5	0.409	-60.9	0.521	-44.9	0.636

Notes: N= 408 in the nonprobability sample, N=253 in the probability sample. Weight 1 is based on socio-demographic variables only, weight 2 includes financial behaviours, and weight 3 includes digital affinity variables. Delta = percentage point difference between the nonprobability sample (not shown) and the probability sample estimate in column 1. In the upper panel, percent of respondents reporting

£0 expenditure in the given category, P = p-values from χ^2 tests for differences in distributions between the two samples. In the lower panel, mean of non-zero expenditure reports, P = p-value from tests of differences in means between the two samples.